Audio Signal Processing : VII. Time-frequency analysis

Emmanuel Bacry

bacry@ceremade.dauphine.fr
http://www.cmap.polytechnique.fr/~bacry

Definition :

$$G(t,\omega) = \int s(u)g(u-t)e^{-i\omega u}du$$

where

- $s(t)$: is the audio signal
- $g(t)$: is the window (localized, symetric, real function)

What shape for the window ?

m g should be smooth enough not to introduce spurious high frequency (more to come later)

$$G(t,\omega) = \int s(u)g(u-t)e^{-i\omega u}du$$

It can be rewritten

$$G(t,\omega) = < s, g_{\omega,t} >$$

with

$$g_{\omega,t}(u) = g(u-t)e^{i\omega u}$$

$g_{\omega,t} \simeq$ time-frequency atoms $\simeq$ "test" functions

- Localization in time
  - Centered at time t
  - support $\simeq \sigma_t = \Delta t$
- Localization in frequency $\hat{g}_{\omega,t}(\xi) = e^{-i\omega t}\hat{g}(\xi - \omega)$
  - Centered at frequency ω
  - support $\simeq \sigma_\omega = \Delta\omega$

Heisenberg inequality : $\Delta t \Delta\omega \geq C$

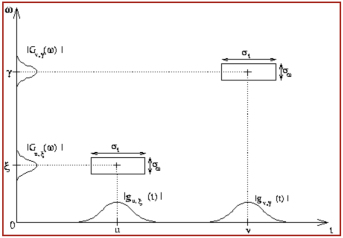
$$G(t,\omega) = \int s(u) g(u-t) e^{-i\omega u} du$$

It can be rewritten

$$G(t,\omega) = < s, g_{\omega,t} >$$

with

$$g_{\omega,t}(u) = g(u-t) e^{i\omega u}$$

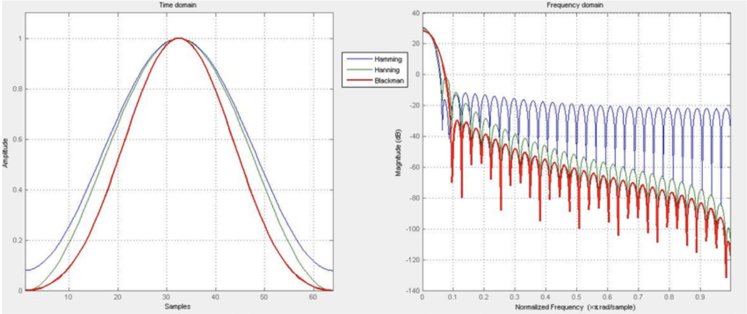


What shape for the window ?

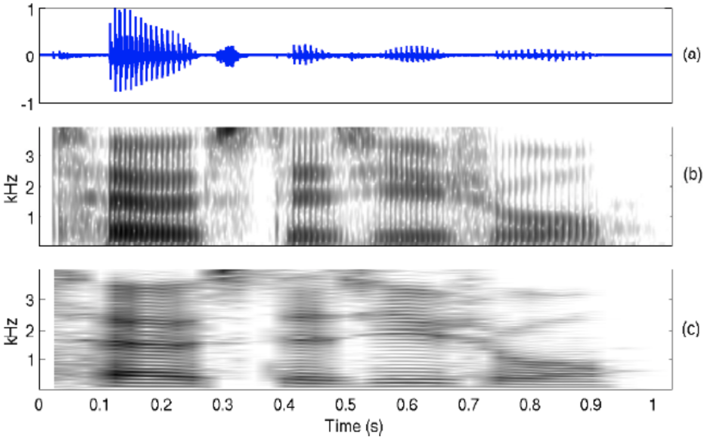
- Hanning : $\frac{1}{2} - \frac{1}{2}\cos(2\pi x),\ x \in [0,1]$
- Hamming : $0.54 - 0.46\cos(2\pi x),\ x \in [0,1]$
- Blackman : $0.42 - 0.5\cos(2\pi x) + 0.08\cos(4\pi x),\ x \in [0,1]$

Fenêtre	Lobe 2aire (dB)	Pente (dB/oct)	Bande passante (bins)	Perte au pire des cas (dB)
Rectangulaire	–13	–6	1,21	3,92
Triangulaire	–27	–12	1,78	3,07
Hann	–32	–18	2,00	3,18
Hamming	–43	–6	1,81	3,10
Blackman-Harris 3	–67	–6	1,81	3,45

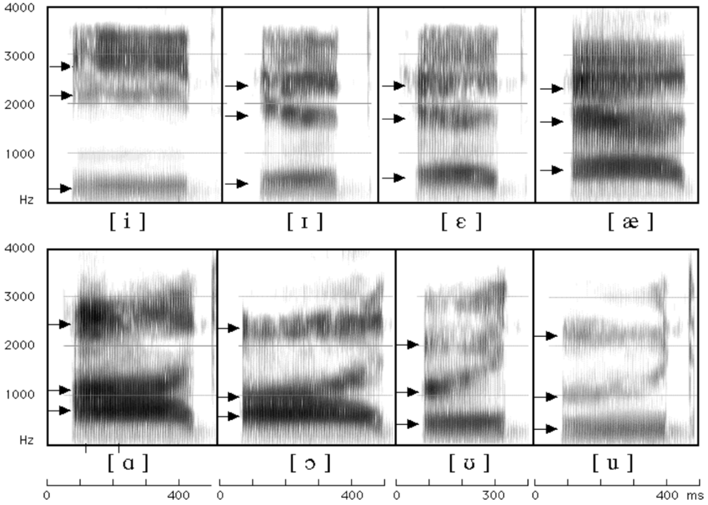
What shape for the window ?



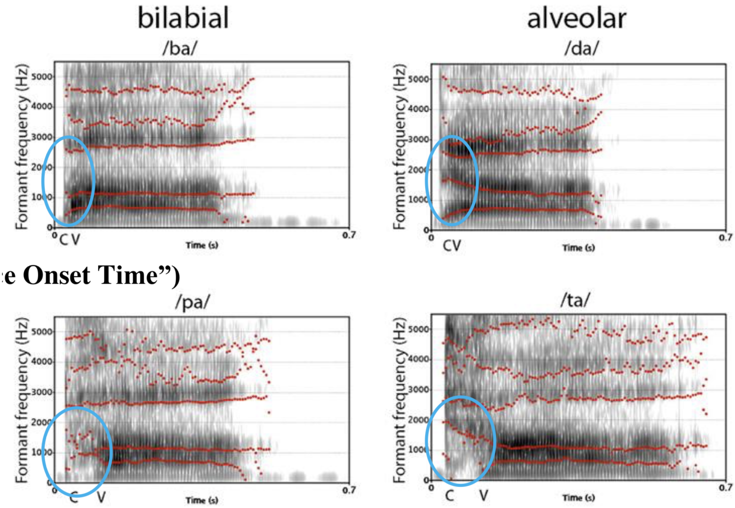
Example of Spectrograms : narrow versus wide-band



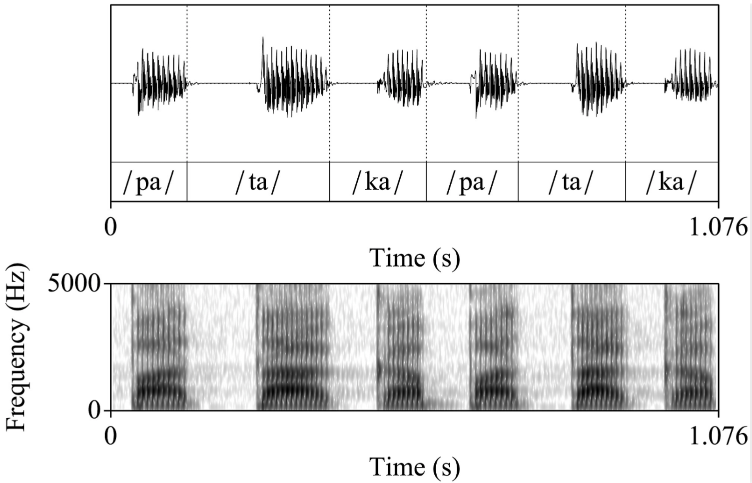
Example of Spectrograms : vowels



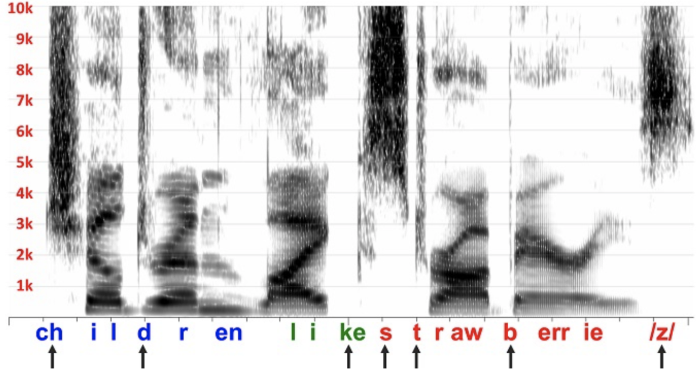
Example of Spectrograms : plosives



Example of Spectrograms : more plosives



Example of Spectrograms : sentence



Definition

$$G(t,\omega) = \int s(u)g(u-t)e^{-i\omega u}du$$

One gets

$$G(t,\omega)e^{i\omega t} = s \star g_\omega(t)$$

where $g_\omega(t) = g(t)e^{i\omega t}$, $\hat{g}_\omega(\xi) = \hat{g}(\xi - \omega)$

Since $G(t,\omega) =< s, g_{\omega,t} >$, with $g_{\omega,t}(u) = g(u-t)e^{i\omega u}$

We could expect a reconstruction formula like

$$s(t) = C \int d\omega \int du \; G(t,\omega) g_{\omega,t} \quad ?$$

Actually one can use a reconstruction window h different from the analysis window g

$$s(t) = C \int d\omega \int du \; G(t, \omega) h_{\omega,t}$$

where

$$C = \frac{1}{2\pi < h, g >}$$

Definition

$$G(t,\omega) = \int s(u)g(u-t)e^{-i\omega u}du$$

One has

$$||s||^2 = \frac{1}{2\pi||g||^2}\int du \int d\omega \; |G(u,\omega)|^2$$

Analysis

- Time discretization (Shannon) of the signal $s[n] = s(n\Delta t)$

$$G(\omega, n\Delta t) = \sum s[m]g[m-n]e^{-i\omega m\Delta t}$$

- g has a support of size N (so does $s[.]g[.-n]$)
- A natural sampling for ω is given by Discrete Fourier Transform : $\Delta\omega = \frac{2\pi}{N\Delta t}$

$$G[k,n] = G(k\Delta\omega, n\Delta t) = \sum_{m=0}^{N-1} s[m]g[m-n]e^{-\frac{2i\pi km}{N}}, \quad n \in [0, N[, \quad k \in [0, N$$

Any hint about increasing the sampling precision of ω (i.e., decreasing $\Delta\omega$)?

Reconstruction

- We subsample in time $\{G(k,n)\}_{k,n} \to \{G(k,pR)\}_{k,p}$
- For each time pR we inverse Fourier transform, so we get $\{s[n]g[n-pR]\}$ (which has a support of size N).
- We use a reconstruction window $h[n]$ such that

$$\sum_p g[n-pR]h[n-pR] = 1$$

- we thus get

$$\sum_p s[n]g[n-pR]h[n-pR] = s[n]$$

Framework

$$s(t) = a(t)\cos(\phi(t))$$

where

- $a(t)$: slowly varying (compared to $\phi(t)$)
- $\phi'(t)$: instantaneous frequency
- $\phi'(t)$: slowly varying (compared to $\phi(t)$)

Theorem

$$s(t) = a(t)\cos(\phi(t)), \quad \Delta a(t) << \Delta\phi(t), \quad \Delta\phi'(t) << \Delta\phi(t)$$

If $\hat{g}(\omega)$ has a support $]-\frac{\Delta\Omega}{2}, \frac{\Delta\Omega}{2}[$ and if $\phi'(t) > \frac{\Delta\Omega}{2}$ then the function

$$\omega \longrightarrow |G(t,\omega)|$$

has a maximum at $\omega = \phi'(t)$